

TRABAJO DE FINAL DE GRADO EN TRADUCCIÓN E INTERPRETACIÓN

TREBALL DE FI DE GRAU EN TRADUCCIÓ I INTERPRETACIÓ

Departament de Traducció i Comunicació

TÍTULO / TÍTOL

La representación de las personas refugiadas en el inglés original e interpretado de los discursos parlamentarios europeos: una aproximación desde los Estudios de Traducción e Interpretación de Corpus (ETIC)

Autor/a: José Aleixandre Becerra

Tutor/a: María Calzada Pérez

Fecha de lectura/ Data de lectura: junio de 2018



Resumen/ Resum:

(aprox. 300 palabras / paraules)

En este trabajo se pretende explorar la representación de las personas refugiadas en el inglés original e interpretado de las intervenciones parlamentarias europeas de los meses de febrero y marzo de 2016, con tal de lograr una mejor comprensión de la situación de crisis prolongada en la que se halla la UE y de hallar respuesta a las preguntas que nos planteamos mediante las tecnologías y metodologías de los Estudios de Traducción e Interpretación de Corpus (ETIC).

Tras una contextualización y breve explicación teórica sobre el contexto espacio-temporal de la comunidad discursiva que analizamos, se ha procedido a una compilación (o transcripción) y a un etiquetado documental semiautomatizado para el posterior análisis, con WordSmith Tools 6.0, de listas de frecuencia, palabras clave, colocaciones y agrupaciones. Así, el contraste de originales en inglés e interpretaciones al inglés nos ha permitido extraer conclusiones sobre el estado de la UE y el contexto de parálisis y confusión en el que se encuentra como consecuencia, principalmente, de su incapacidad de gestionar la diversidad.

Palabras clave/ Paraules clau: (5)

ETIC, personas refugiadas, PE, análisis del discurso, lingüística de corpus

Estilo usado en las citas y referencias bibliográficas: estilo APA

Para vosotrxs, porque os traicionaron. Para lxs nadie.

أبها المارون في الكلمات العابرة
احملوا أسمائكم وانصرفوا
وأسحبوا ساعاتكم من وقتنا، وانصرفوا
وخذوا ما شئتم من زرقاة البحر و رمل الذاكرة
وخذوا ما شئتم من صور، كي تعرفوا
انكم لن تعرفوا
كيف يبني حجر من ارضنا سقف السماء

Pasajeros entre palabras fugaces
Mahmud Darwish

*Les frontières, sont-elles le dernier bastion de nos identités
ou le dernier rempart de nos espoirs ?*

Ulysse from Bagdad
Éric-Emmanuel Schmitt

Índice

Índice de figuras y tablas	6
1. Introducción y contextualización.....	7
2. Marco teórico.....	9
2.2. El Parlamento Europeo (PE) como «comunidad discursiva». El proyecto ECPC.....	9
2.3. La lingüística de corpus (LC): introducción y tipología.....	10
2.3.1. Tipologías de corpus	11
2.4. Los Estudios de Traducción e Interpretación de Corpus (ETIC).....	12
2.5. Los corpus: usos y aplicaciones	13
3. Metodología.....	14
3.1. Enfoques metodológicos de los Estudios de Traducción e Interpretación de Corpus	14
3.2. Compilación del corpus	14
3.2.1. Preguntas de partida y condicionantes	14
3.2.2. Composición del corpus	16
3.2.3. Extracción y transcripción de los discursos	16
3.3. Etiquetado y enriquecimiento automatizado de los discursos con metadatos	17
3.4. Herramientas de análisis: WordSmith Tools 6.0.	19
4. Análisis del corpus.....	20
4.1. Listas de frecuencia (<i>wordlist</i>).....	20
4.2. Palabras clave (<i>keywords</i>).....	23
4.3. Concordancias (<i>concord</i>)	25
4.4. Colocaciones (<i>collocations</i>) y agrupaciones (<i>clusters</i>).....	29
5. Conclusiones.....	31
Bibliografía.....	33

Índice de figuras y tablas

Figura 1. Clasificación de los corpus lingüísticos (Rabadán y Fernández Nistal, 2002)	11
Figura 2. Ejemplo de discurso original en inglés sin etiquetar en .txt.....	18
Figura 3. Ejemplos de expresiones regulares para el etiquetado automático	18
Figura 4. Ejemplo de discursos etiquetados documentalente en XML	19
Figura 5. Datos estadísticos sobre el corpus (Fuente: elaboración propia)	21
Figura 6. Wordlist de los originales.....	22
Figura 7. Wordlist de las interpretaciones	22
Figura 8. Lista de <i>positive keywords</i>	24
Figura 9. Lista de <i>negative keywords</i>	24
Figura 10. Enlaces entre palabras clave.....	25
Figura 11. Concordancias obtenidas a partir de la búsqueda "refugee?" en el corpus de interpretaciones	27
Figura 12. Concordancia a partir de la búsqueda "immigrant?" en el corpus	28
Figura 13. Concordancia obtenida a partir de la búsqueda "seeker?"	29
Figura 14. Agrupaciones a partir de <i>crisis</i>	30

1. Introducción y contextualización¹

Tras la llegada a Europa de miles de personas como consecuencia de los conflictos que asolan algunos países como Siria, Somalia, Yemen, Iraq o Libia, entre otros, la población europea se enfrenta a algo para lo que nadie la ha preparado; se esfuerza en poner nombre a lo desconocido sin muchas veces llegar a entenderlo. En esta «era de la nostalgia» (Bauman, 2017), que es como define el filósofo Zygmunt Bauman nuestro tiempo, se habla de la «crisis de los refugiados» y del «pánico migratorio», que es fruto, según él, del

desconocimiento de cómo continuar, de cómo tratar una situación que no hemos creado y que no tenemos bajo control, [lo que] causa gran ansiedad y miedo (Bauman, 2016, p.15).

Si bien es cierto que siempre ha habido personas que se han visto obligadas a abandonar sus países para refugiarse de las guerras o de una existencia sin futuro, nunca antes la humanidad ha sentido tanta incomodidad y parálisis ante, por ejemplo, una imagen que dio la vuelta al mundo y evidenció el fracaso de la UE para gestionar esta situación: la del ahogamiento, en septiembre de 2015, frente a la costa turca, de Aylan Kurdi, un niño refugiado sirio de 3 años.

En octubre de 2015 la situación se volvió extrema: 10.000 llegadas diarias a las costas griegas². Entre febrero y marzo de 2016, período que analizamos en nuestro corpus, el Parlamento debate sobre el que fue bautizado por algunos medios como el «pacto de la vergüenza»³ entre la UE y Turquía, que el Consejo Europeo (CE) acabó aprobando el 18 de marzo de 2016 con el objetivo de «poner término a la migración irregular» (CE, 2016). Mediante el desembolso de 3000 millones de euros, la UE aprobaba la devolución de miles de personas a Turquía, alegando los «esfuerzos de seguridad»⁴ del país, precisamente en uno de los episodios de mayor actividad terrorista en el mismo. De este modo, la UE incumplía uno de los principios fundamentales del Derecho internacional, el de no devolución (*non-refoulement*), que recoge el artículo 33 de la Convención Sobre el Estatuto de los Refugiados de 1951⁵.

La elección de este tema no ha sido casual, ya que lo que se busca es explorar un fenómeno tan complejo como éste sirviéndonos de las posibilidades de análisis que nos ofrecen los Estudios de Traducción e Interpretación de Corpus (ETIC) (en inglés: *Corpus-based Translation and Interpreting Studies*, CTIS), con el objetivo de lograr una mejor comprensión del mismo.

Nuestro trabajo se ha organizado en tres partes: un marco teórico, para introducir el contexto de la comunidad discursiva en la que nos encontramos así como para explicar el desarrollo de los ETIC en

¹ Este trabajo se ha producido en el marco del proyecto *Estudia i Investiga a l'UJI*, gracias al cual he colaborado con el grupo de investigación European Comparable and Parallel Corpus (ECPC) y con con los proyectos: ECPC+: (Generalitat Valenciana, AICO/2017/082) y MD-ECPC (de la Universitat Jaume I, UJI-B2017-25).

² Según un informe del Consejo Europeo del 17/03/2017. Disponible en: https://ec.europa.eu/homeaffairs/sites/homeaffairs/files/what-we-do/policies/european-agenda-migration/background-information/eu_turkey_statement_17032017_es.pdf [Consulta: 09/02/2018].

³ Véase, por ejemplo: https://elpais.com/elpais/2016/03/08/opinion/1457424011_548168.html [Consulta: 09/02/2018].

⁴ Comunicado de prensa del 18/03/2016, disponible en: <http://www.consilium.europa.eu/es/press/press-releases/2016/03/18/eu-turkey-statement/> [Consulta: 09/02/2018].

⁵ Véase: http://www.acnur.es/PDF/0005_20120511114519.pdf [Consulta: 09/02/2018].

tanto que paradigma surgido desde los estudios descriptivos de la traducción, la lingüística de corpus y el análisis crítico del discurso. Un segundo apartado en el que, dadas las marcadas características metodológicas de los ETIC, explicaremos el proceso de compilación y etiquetado de nuestro corpus, así como las herramientas utilizadas en dicho proceso como en el de análisis. Por último, en un tercer apartado, presentaremos los resultados más significativos obtenidos en nuestro análisis para llegar a conclusiones que respondan o refuten algunas nociones teóricas apriorísticas con las que exploramos este fenómeno.

2. Marco teórico

Muchas son las disciplinas que desde hace tiempo se vienen interesando por el análisis de la representación discursiva de las personas refugiadas en busca de una mejor comprensión de la misma, y la lingüística de corpus (LC) y el análisis crítico del discurso (CDA por sus siglas en inglés y, en adelante, ACD) no han sido la excepción; lo que corroboran los numerosos trabajos dedicados al tema desde una u otra disciplina o bien desde una combinación productiva de ambas (P. Baker, 2006; P. Baker et al., 2008; Gabrielatos y Baker, 2008; Van Dijk, 1995), aunque casi siempre desde el análisis del discurso, entendido como práctica social (Fairclough, 1993, p.63), de la prensa escrita.

La metodología de corpus, combinada con la de análisis del discurso y los estudios descriptivos de traducción (DTS por sus siglas en inglés, y, en adelante, EDT), es la raíz de los ETIC, por lo que resulta de interés (así lo demuestran trabajos como el de Paul Baker (2006) abordar este fenómeno desde una perspectiva mixta (la parte cuantitativa de la LC y la cualitativa del ACD). Y es que «refugees are a particularly interesting subject to analyse because they consist of the most relatively powerless groups in society» (*ibíd.*: p.73-74).

Por ello, pasamos a continuación a desarrollar algunos de los conceptos básicos que vertebran este trabajo con tal de proporcionar una justificación teórica al análisis cualitativo y cuantitativo que realizaremos en el apartado 4, lo que pasa por comprender el contexto en el que se desarrollan las intervenciones que analizamos, el PE, y, en último término, la relación de los EDT, la LC y el ACD como base de los ETIC, cuyos postulados nos servirán de punto de partida para nuestro análisis.

2.2. El Parlamento Europeo (PE) como «comunidad discursiva». El proyecto ECPC

Con el objetivo de lograr una mejor comprensión de los discursos parlamentarios europeos resulta esencial entender, en primer lugar, el contexto en el que se desarrollan: el PE; una cámara legislativa multinacional y multilingüe, formada por 751 diputados y diputadas, cuya elección se lleva a cabo por medio de sufragio universal directo, que representan a 500 millones de ciudadanos y ciudadanas de 28 estados miembros y con sede en Bruselas y Estrasburgo, siendo esta última ciudad donde se llevan a cabo los debates parlamentarios una vez al mes⁶.

Y es que, como apunta Calzada Pérez, la importancia del PE en el contexto global actual radica en que

[...] much of Europe's most important decision making occurs in the Euro Chamber through multilingual exchanges, and its agenda influences ideologies, behaviours and the language of national chambers (2017, p.278).

Calzada Pérez (2007) defiende que existe una interdependencia entre texto y contexto, por lo que

⁶ Véase: <http://www.europarl.europa.eu/aboutparliament/es> [Consulta: 09/02/2018].

se hace necesario un análisis macroestructural de los discursos del PE antes de proceder a un análisis textural siguiendo la metodología de los ETIC. Y aunque este análisis macroestructural del género del debate parlamentario se encuentra fuera de los objetivos del presente trabajo, consideramos relevante recordar que las intervenciones de los parlamentarios europeos (tanto orales como escritas) pueden enmarcarse dentro de aquello que se ha denominado «comunidad discursiva» (Swales, 1990, p.24-27), lo que se traduce en el uso de una jerga específica dentro del PE (eurojerga) (cf. Koskinen, 2011), en el hecho de que el PE es una comunidad con diferentes grados y tipos de participación de sus componentes, y en la existencia de géneros textuales así como de un comportamiento discursivo propios, ya que las intervenciones parlamentarias, en este caso, «son el canal de transmisión de la ideología —compleja e híbrida—, de la comunidad discursiva del PE» (Calzada Pérez, 2012, p.36-40).

Resulta imposible desvincular la trayectoria del grupo ECPC (*European Comparable and Parallel Corpora*)⁷ del PE. Dirigido por María Calzada Pérez, en ECPC han colaborado y colaboran algunos de los estudiosos de más prestigio en el ámbito de los ETIC, como la Dra. Dorothy Kenny, la Dra. Mona Baker, la Dra. Silvia Bernardini o el Dr. Jeremy Munday, entre otros, y se fija como objetivo principal, en palabras de Martínez y Serrat (2012, p.66), «conocer y profundizar en el estudio del discurso parlamentario como género textual con especial atención a la influencia de la traducción en dicho género».

ECPC constituye, así pues, un ejemplo paradigmático de las posibilidades de análisis y explotación de corpus electrónicos, comparables y paralelos, desde los ETIC, mediante tecnologías especialmente punteras y aprovechando el potencial del uso de corpus para la traducción y la interpretación, que detallaremos en los siguientes subapartados.

2.3. La lingüística de corpus (LC): introducción y tipología

En palabras de McEnery y Hardie (2012, p.1), la LC no se encarga del estudio de un aspecto concreto del lenguaje sino que «it is an area which focuses upon a set of procedures, or methods, for studying language». De este modo, según Biber et al. (1998, p.4), la LC presenta las siguientes características: es empírica y analiza el lenguaje tal como lo emplean los usuarios en contextos concretos; utiliza una colección de textos, un corpus, como base de su análisis; se sirve de la tecnología y aplica técnicas automáticas e interactivas; y su análisis depende de parámetros mixtos. Por lo tanto, no es baladí decir que la metodología es lo que ha acabado definiendo la disciplina (Tognini Bonelli, 2010, p.18) y que, en el desarrollo de la LC, la evolución tecnológica ha desempeñado una función decisiva.

Así pues, y teniendo en cuenta que existen casi tantas definiciones de corpus como lingüistas que los estudian, nos remitiremos a uno de los padres de la LC, John Sinclair, que lo define como «a collection of naturally-occurring language text, chosen to characterize a state or variety of a language»

⁷ Sobre ECPC: <http://www.ecpc.uji.es/> [Consulta: 10/02/2018].

(Sinclair, 1991, p.171), a lo que cabría añadir sus posibles aplicaciones para la enseñanza-aprendizaje de lenguas, la lingüística forense, la pragmática, la sociolingüística, el análisis del discurso (mediático y político) y, por supuesto, la traducción y la interpretación (O’Keeffe y McCarthy, 2010, p.7-12).

2.3.1. Tipologías de corpus

Casi todas las investigadoras e investigadores que hablan de LC han propuesto un modelo de clasificación de corpus (Biber, 1994; Granger, 2003; Rabadán y Fernández Nistal, 2002; Roberts, 2006; Tognini Bonelli, 2010; Kenny, 2001; Laviosa, 2002; Olohan, 2004; Tognini-Bonelli y Sinclair, 2006), pero hoy por hoy no existe consenso terminológico ni clasificatorio, por lo que se hace necesario una revisión de las distintas tipologías.

En un primer nivel clasificatorio, Roberts (2006) distingue entre corpus generales y corpus lingüísticos, siendo estos últimos los que abordan cuestiones lingüísticas. A partir de este nivel, dentro de la categoría de corpus lingüísticos, Rabadán y Fernández Nistal (2002) trazan el siguiente esquema de clasificación:

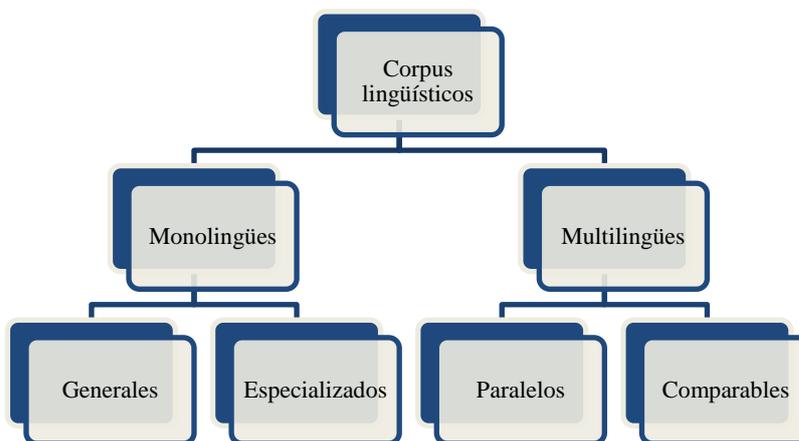


Figura 1. Clasificación de los corpus lingüísticos (Rabadán y Fernández Nistal, 2002)

Los corpus monolingües (textos pertenecientes a una sola lengua) pueden ser generales o especializados. Los bilingües y/o multilingües están compuestos por fragmentos en varios idiomas, ya sean originales acompañados de traducción (paralelos) o bien originales en varias lenguas que, aun no siendo traducciones, presentan semejanzas y están relacionados (comparables). Estos dos últimos, como señalan McEnery y Xiao (2008, p.21), se usan sobre todo en traducción y estudios contrastivos, aunque, como apuntan más tarde (*ibíd.*: p.27) el uso de los corpus comparables en traducción «provides a poor basis for cross-linguistic contrast if used as the sole source of data», por lo que lo ideal es contar con corpus tanto paralelos como comparables, que, sin embargo, escasean (Kenning, 2010, p.488).

A partir de ahí los límites se vuelven difusos y surgen diferentes clasificaciones y aportaciones que nutren esta clasificación primaria. Así, Teubert y Čermáková (2004) y McEnery y Xiao (2008) hablan de corpus oportunistas o caníbales, resultado de compilar «all the corpora one can lay hands upon» (Teubert y Čermáková, 2004, p.120). A este tipo de corpus cabe añadir el corpus abierto (o

monitor corpus), que se mantiene en constante actualización y es indicador de, por ejemplo, los cambios en la lengua.

Por su parte, Kenny (2001) y Olohan (2004) proponen otra división dentro de los corpus monolingües, entre *single monolingual corpora* (originales en una sola lengua) y *single translational corpora* (traducidos en una sola lengua). Un ejemplo de este último es el Translational English Corpus (TEC) que, tal como señala Calzada Pérez (2012, p.71), es una de las bases más firmes de los ETIC e inspira el proyecto ECPC.

Antes de abundar en los ETIC, cabe mencionar, por último, la propuesta de Granger (2003, p.21), que introduce un giro en la tipología de corpus. Para Granger, los corpus pueden ser monolingües o multilingües. Dentro de estos últimos distingue entre comparables y paralelos/traductores. Para los primeros propone una bifurcación entre originales y traducciones y para los segundos entre unidireccionales y bidireccionales. En los monolingües, Granger sólo habla de comparables que, sin embargo, pueden ser de originales y traducciones (TO+TT) o de textos nativos y no nativos. A modo de conclusión se puede afirmar, en palabras de Calzada Pérez (2012, p.73), que esta clasificación «establece espacios flexibles y está especialmente adaptada a la traducción».

2.4. Los Estudios de Traducción e Interpretación de Corpus (ETIC)

El uso de corpus, al menos en traducción, no supone ya ninguna novedad. Hace tiempo que quienes traducen y quienes estudian la traducción se vienen interesando por las posibles aplicaciones de los corpus en nuestra disciplina, en especial como poderoso recurso de documentación (Kübler y Aston, 2010, p.502). De hecho, como precisa Laviosa (2002, p.33), cada vez es mayor el número de personas que creen en el potencial del corpus como vehículo de reconciliación entre la pluralidad de intereses y las necesidades de la disciplina.

El origen de los ETIC hemos de buscarlo principalmente, además de en la LC, en los estudios descriptivos de la traducción, cuyo principal exponente es Gideon Toury. En sus orígenes, los EDT se planteaban, entre otras cosas, formular normas traductoras (cf. Toury (1995) que consiguieran acabar con el prescriptivismo anterior, aunque con las limitaciones propias de la explotación manual de los corpus examinados. Sin embargo, Mona Baker fue la primera en aprovechar el potencial de los corpus en traducción, lo que la llevó a crear uno de los primeros corpus traductores del mundo: el *Translational English Corpus* (TEC), con traducciones de textos de ficción y no ficción de varias lenguas hacia el inglés. Baker se aventuró incluso a formular la hipótesis de los «universales de la traducción» (1993, p.243), una serie de elementos que habitualmente aparecerían en las traducciones, con independencia del par de lenguas en cuestión, y que serían una característica distintiva de los textos traducidos respecto de los originales.

Así pues, esta base teórica no sólo inaugura una nueva metodología sino que, para algunas (Laviosa, 2002, p.22), inaugura un nuevo paradigma en nuestra disciplina:

[...] The growth and variety of research is such that we can no longer regard Corpus-based Translation Studies merely as an innovative methodology or a new perspective. It is now a coherent, composite, and rich paradigm that addresses fundamental theoretical, descriptive and practical issues within the discipline.

2.5. Los corpus: usos y aplicaciones

Los corpus no sólo participan en el conjunto de tareas que conforman el proceso de traducción e interpretación, especialmente en la fase de documentación (Kübler y Aston, 2010, p.501), sino que poseen múltiples aplicaciones más, si bien, de todas ellas, la que quizá se hace más patente es la del uso del corpus como fuente documental y terminológica, tanto para traductoras e intérpretes en formación como para profesionales (Bowker, 1998; Kenning, 2010; Calzada Pérez, 2012; Laviosa, 2002; Ajmer y Altenberg, 1996), pues facilita la resolución de problemas traductológicos, ayuda a producir traducciones de mayor calidad, profesionalidad y consistencia o, sencillamente, fomenta una mayor sensibilidad lingüística y cultural (Bernardini, 1997).

La segunda gran aplicación de los corpus es en la enseñanza-aprendizaje de la traducción y la interpretación, por ejemplo para mejorar la comprensión del T.O. y producir T.T. más fluidos, al igual que para traducir textos de alto contenido especializado (cf. Zanettin, 1998; Friedbichler y Friedbichler, 1997; Gavioli, 1997, Bowker, 1998; Bowker, 2000), con tal de mejorar la productividad y cometer menos errores, incluso para la enseñanza de las lenguas de trabajo y, en general, la enseñanza-aprendizaje de lenguas (Barlow, 2000). En el caso de las intérpretes, además, como señalan investigaciones recientes, el uso de corpus puede convertirse en una herramienta de adquisición de «la lengua de la interpretación» (Aston, 2018).

Por último, y aunque existen muchas más aplicaciones, como el uso de corpus para el desarrollo de los sistemas de TA (Somers, 2003; Hutchins, 2003), la utilización de corpus es una herramienta de gran utilidad en la investigación en traducción e interpretación, y prueba de ello son los numerosos grupos de investigación que utilizan corpus con fines diversos, como es el caso del grupo ECPC, el grupo GENTT o el grupo COVALT, por citar algunos de nuestra propia institución.

3. Metodología

3.1. Enfoques metodológicos de los Estudios de Traducción e Interpretación de Corpus

Tal como comentábamos en el apartado anterior, los ETIC son un paradigma en cuyo desarrollo ha influido tanto una metodología característica como el desarrollo tecnológico. Quienes investigan con corpus suelen hablar a menudo de la existencia dos enfoques metodológicos, *corpus-based* y *corpus-driven*, que estarán marcados por una pregunta de investigación articulada (Reppen, 2010, p.31) y unos objetivos concretos. En el primer enfoque, más deductivo, la teoría es anterior a los datos y los datos sirven para validar o refutar la teoría (Marco, 2013, p.10-11). En el segundo, sin embargo, más inductivo, se formula una teoría a partir de los datos obtenidos (Calzada Pérez, 2012, p.95).

En nuestro caso, hemos procurado apoyarnos en un enfoque *corpus-driven*, en parte porque es mediante este enfoque, que consideramos más abierto a los datos y menos dependiente de nociones teóricas apriorísticas, como se puede lograr, a nuestro modo de ver, una mejor comprensión de un fenómeno tan complejo y poliédrico como el que analizamos. No obstante, dejar emerger las conclusiones a partir de los datos no es incompatible con tener ideas y conocimientos previos acerca del tema. De hecho, en palabras de Baker (2006, p.178), los esfuerzos de creación de un corpus de poco sirven sin alguien que los interprete y los interprete con todos los conocimientos que posee.

Otros dos enfoques que suelen generar polémica son el cuantitativo y el cualitativo o, dicho de otro modo, la dialéctica representada especialmente por la lingüística de corpus y el análisis crítico del discurso. Así, Tymoczko (1998, p.658) ya alertaba de los peligros del exceso de cuantificación en el estudio de la traducción, que es una disciplina variable y que se halla sometida a múltiples factores. Por su parte, conscientes de las limitaciones de un enfoque meramente cuantitativo para analizar el discurso, Gabrielatos y Baker (2008, p.8) se proponen reconciliar dos visiones diferentes —la cuantitativa de la LC y la cualitativa del ACD—, y demostrar que un enfoque mixto puede ser muy productivo para ambas. Siguiendo la estela de, entre otros, estos investigadores, nos hemos servido de un enfoque mixto que aporte riqueza y luz a nuestro análisis.

3.2. Compilación del corpus

3.2.1. Preguntas de partida y condicionantes

No hay duda de que la recopilación de textos para su posterior explotación y análisis es un proceso que requiere, tal como afirman Rabadán y Fernández Nistal, «tiempo, esfuerzo y recursos humanos» (2002, p.56). Y, aunque quien compila corpus para su posterior análisis goza de gran libertad (aun con algunos condicionantes) sobre cómo hacerlo (Baker, 2006, p.45), resulta «imprescindible abordar toda una serie de cuestiones ineludibles», con tal de no fracasar (Rabadán y Fernández Nistal, 2002 *apud* Calzada Pérez, 2012, p.84).

De acuerdo con Reppen (2010, p.32), la primera pregunta que debemos plantearnos antes de comenzar es la del tamaño de nuestro corpus. Como ya señalan otros, por ejemplo Paul Baker (2006), la del tamaño no es una cuestión sencilla para la que exista respuesta o consenso; por un lado, está relacionada con la representatividad, esto es, si disponemos de la cantidad suficiente de textos para representar el lenguaje de la comunidad discursiva que vamos a analizar; y por otro, con la practicidad: ¿tenemos el tiempo suficiente? De igual modo, hemos de considerar algunas cuestiones de carácter práctico, tales como los derechos de autor de los textos que vamos a utilizar (Calzada Pérez, 2012, p.87). Este último aspecto no ha supuesto ningún problema, pues el PE publica los textos originales y los vídeos con interpretaciones en la web, públicamente, para que cualquier persona pueda consultarlos.

En el diseño del corpus también han influido criterios espacio-temporales, pues la elección del período de tiempo no ha sido casual, sino que queríamos ver cómo se había tratado la situación de las personas refugiadas en el PE antes de que el Consejo decidiera aprobar un acuerdo con Turquía para su “devolución”. Ese acuerdo se aprobó finalmente el 18 de marzo, por lo que resultaba interesante compilar textos de dicho mes y del mes anterior, puesto que

[...] the issues surrounding the social conditions of production and interpretation of texts are important in helping the researcher understand discourses surrounding them (Fairclough, 1989 *apud* Baker, 2006, p.18).

Tampoco es fácil determinar la representatividad de nuestro corpus únicamente a partir del número de palabras. Es evidente que, cuanto mayor sea el número de textos que contenga el corpus, más representativo será. Sin embargo, dado que el nuestro pretende ser un corpus especializado, ya que emana de una comunidad discursiva concreta, hemos tenido presentes las palabras Tribble (1997), Baker (2006, p.28) y McEnery y Hardie (2012, p.10), en la medida en que no necesitábamos construir un corpus de millones de palabras, pues el género en el que se inscriben nuestros discursos está ya delimitado de algún modo.

Por otra parte, uno de los mayores condicionantes para la construcción del corpus, también de tipo espacio-temporal, ha sido la ausencia de traducciones para el período estudiado. En julio de 2011, el PE decide dejar de traducir las intervenciones de los plenarios y los sustituye por vídeos (sin subtítulos) y una versión arcoíris; esto es, los discursos en las lenguas en las que originalmente fueron pronunciados pero sin traducir. Este hecho no sólo enlentece la tarea de compilar corpus de discursos parlamentarios europeos, sino que, como alerta Calzada Pérez (2017, p.283) se trata de una forma encubierta y barata de eliminar la transparencia, la accesibilidad, de silenciar al otro; cosa que pasa factura al «European sense of identity». No obstante, de acuerdo con Calzada Pérez (*ibíd.*: p.277), este hecho nos obliga a replantearnos otro tipo de aproximaciones y de análisis, lo que, quizá, podría abarcar la creación de más corpus orales que, según Adolphs y Knight (2010, p.48), constituyen un recurso de gran valor para el estudio del discurso porque se desarrollan marcos multimodales, a pesar de que su construcción conlleva

una cantidad mayor de trabajo a causa de «the burden of transcription» (*ibíd.*: p.49) y de que, hasta ahora, lo escrito ha sido predominante⁸.

3.2.2. Composición del corpus

Con todo ello, hemos creado un corpus *ad hoc*, que hemos denominado ECPC-ETIC-refugees y que cuenta con una extensión aproximada de 370.000 palabras. Nuestro corpus se compone, a su vez, de dos subcorpus: uno de intervenciones originales en inglés y otro de transcripciones de las interpretaciones simultáneas al inglés desde otras lenguas. Siguiendo la clasificación propuesta por Sylvianne Granger (2003), podemos clasificarlo como monolingüe, comparable y especializado, en la medida en que se compone de intervenciones originales en inglés y transcripciones de interpretaciones simultáneas de otras lenguas hacia al inglés procedentes de la misma comunidad discursiva.

El corpus de originales, por un lado, recoge todas las intervenciones en inglés original de oradoras y oradores que tienen el inglés como lengua materna, o bien que se expresan en inglés, que se corresponden con las sesiones plenarias de Estrasburgo de los días 1, 2, 3, 4, 24 y 25 de febrero y del 7, 8, 9 y 10 de marzo de 2016.

El corpus de interpretaciones, por otro lado, recoge las transcripciones de las interpretaciones simultáneas al inglés desde otras lenguas pertenecientes al punto diez de la sesión plenaria del 2 de febrero de 2016 (con título: “Crisis de los refugiados, control de las fronteras exteriores y futuro de Schengen - Respeto del principio internacional de no devolución - Financiación del Mecanismo para Turquía en favor de los Refugiados - Aumento del odio y la violencia racistas contra los refugiados y los migrantes en toda Europa (debate)”) y del punto cinco de la del 8 de marzo de 2016 (con título: “Situación de las mujeres refugiadas y solicitantes de asilo en la UE”).

3.2.3. Extracción y transcripción de los discursos

La extracción de los discursos, al igual que de los vídeos de las interpretaciones simultáneas para su posterior transcripción, se ha llevado a cabo consultando la página del Parlamento Europeo (<http://www.europarl.europa.eu/plenary/es/debates-video.html>), a partir de la cual se han extraído las versiones arcoíris en formato .html, de una parte; y los vídeos para transcribir, de otra.

A partir de ahí, se ha procedido de forma distinta con cada uno de ellos. En el caso de los discursos originales en inglés que figuraban en la versión arcoíris, se ha utilizado un comando de búsqueda para identificarlos en el .html para posteriormente copiarlos y editarlos utilizando principalmente Notepad++, Microsoft Word y jEdit. Todo ello para obtener varios archivos .txt, con una

⁸ Una honrosa excepción es el corpus EPIC, con un corpus multilingüe, de gran calidad, elaborado a partir de las interpretaciones simultáneas de intervenciones del PE, paradigma de los *Corpus-based Interpreting Studies*. Véase: <http://www.dit.unibo.it/it/ricerca/centri-di-ricerca/colitec-corpora-linguistica-tecnologia> [Consulta: 23/04/2018]

nomenclatura diferenciadora, en los que sólo constaran los textos originales en inglés, cuya edición y etiquetado comentaremos en el subapartado siguiente.

En el caso de las interpretaciones, sin embargo, partiendo de la premisa de que la representatividad es más difícil de conseguir porque el proceso de construcción del corpus es más costoso (Strassel y Cole, 2006, p.3 *apud* Adolphs y Knight, 2010, p.41), hemos procurado seleccionar aquellos textos que nos aporten la mayor cantidad posible de contenido y contexto, con tal de obtener resultados más representativos en la fase de análisis.

Ahora bien, la elaboración de un corpus oral se aleja de los objetivos del presente trabajo, en el sentido de que confeccionar un corpus oral no sólo requiere tareas de transcripción, que hoy por hoy no pueden automatizarse, a falta de un mayor desarrollo tecnológico de los sistemas de reconocimiento de voz (*ibíd.*: p.48); sino que también se requiere un etiquetado más completo, puesto que en el discurso oral es fundamental contar con información complementaria. Y, si bien no dudamos de que «the richness of this type of data can make the extra effort worthwhile» (*ibíd.*: p.44), por cuestiones de practicidad hemos tenido que limitarnos a una transcripción y a un etiquetado contextual sencillos.

3.3. Etiquetado y enriquecimiento automatizado de los discursos con metadatos

En palabras de Leech (2004, p.25) «corpus annotation is the practice of adding interpretative linguistic information to a corpus». Tal información interpretativa, que se añade mediante el etiquetado y que puede ser, entre otros, de tipo documental, es, según Sinclair (2004 *apud* Adolphs y Knight, 2010, p.42) «critical to a corpus to help achieve the standards for representativeness, balance and homogeneity».

En este sentido, el etiquetado documental proporciona información contextual esencial de aspectos como, por ejemplo: nombre, grupo político, cargo (si procede), sexo o tipo de discurso (original/interpretación; oral/escrito). Se trata de atribuir un encabezamiento, codificado en .xml, a cada uno de los discursos que conforman el corpus para facilitar su identificación y facilitar posteriormente el análisis con variables como pueden ser las ideológicas o las de sexo.

En nuestro caso, dicho etiquetado se ha llevado a cabo de forma automatizada, mediante el uso de expresiones regulares (RegExp), que son «patrones utilizados para encontrar una determinada combinación de caracteres dentro de una cadena de texto»⁹, con una posterior revisión manual para corregir irregularidades de formato y añadir elementos, como por ejemplo el sexo, de forma manual. En el caso de las transcripciones de las interpretaciones, este proceso ha sido enteramente manual, y los errores se han corregido de forma automatizada.

De este modo, se ha pasado a analizar y descubrir patrones estructurales en los discursos originales, que presentan una apariencia semejante a la siguiente:

⁹ Véase: https://developer.mozilla.org/es/docs/Web/JavaScript/Guide/Regular_Expressions [Consulta: 23/04/2018]

Marianne Thyssen, Member of the Commission . - Madam President, a deeper and a fairer internal market and the fight against social dumping are absolute priorities for
Undeclared work is not only unfair to the workers involved, it is also unfair to businesses that do play by the rules. We know that small and medium-sized companies are
That is why the establishment of a European platform to enhance cooperation in the prevention and deterrence of undeclared work and to tackle undeclared work is such a
It is only through a holistic approach, across borders and across disciplines, that we will be able to tackle this complex phenomenon effectively. I would like to stre
Finally, I note with satisfaction that the social partners will play an important role in the platform and that close cooperation with Eurofound and the European Agency
I would like to thank Parliament for all its efforts in making a timely agreement on the platform possible. In particular, I would like to express my thanks to the rap
I am committed to doing everything possible from my side to make the platform a success. The platform will regularly inform you about its activities and submit its work

Anthea McIntyre, on behalf of the ECR Group . - Madam President, firstly I would like to thank the rapporteur, Mr Pirinski, and all the shadows for their work on this
Undeclared work is a very serious subject which impacts on all the Member States. It particularly affects tax evasion - which is a real problem - benefit fraud, illegal
We need to recognise that the level of undeclared work varies from state to state; for example, in the UK, according to the OECD data, the level of the informal economy

Bill Etheridge (EFDD), blue-card question . - Good morning to my West Midlands colleague. I note that you say it must be done with sovereign countries cooperating and

Anthea McIntyre (ECR), blue-card answer . - In response to the first question, minimum wages have absolutely nothing to do with the platform on undeclared work. I thi
To Bill Etheridge: well of course, if you come from your party, you do not want anything at a European level. I beg to differ. I think that there are cases, and this i

Marian Harkin, on behalf of the ALDE Group . - Madam President, first of all I would like to congratulate the rapporteur Mr Pirinski, and indeed the shadow rapporteurs
The fact that membership is mandatory means that all Member States must participate, and therefore they must be involved in exchanging information and developing exper
The platform will not contradict any national legislation and will not propose harmonisation of any laws, but it will provide a strong framework for enhanced cooperati
In our next debate we look at the anti-tax-avoidance package, which recognises the importance of coordinated action. Equally, this type of coordinated action under the

Figura 2. Ejemplo de discurso original en inglés sin etiquetar en .txt

A continuación, una vez descubiertas las similitudes (e irregularidades) de formato de los discursos y codificado el documento en UTF-8, se ha pasado a etiquetarlos de forma automática mediante el uso de expresiones regulares en el editor jEdit; algunas de ellas, puesto que se han usado muchas más, a título ilustrativo, recogidas en la siguiente tabla:

1. Poner el nombre al principio de cada discurso	Buscar: <code>(\n\r){2}(\p{Lu}.*?)\n</code> Reemplazar: <code>\n\nNAME: \$2\n</code>
2. Añadir el cargo (POSITION) en aquellos discursos en los que sea oportuno	Buscar: <code>^(NAME: .+?),(.*?)\n</code> Reemplazar: <code>\$1\nPOSITION: \$2\n</code>
3. Añadir el sexo del orador o de la oradora (SEX)	Buscar: <code>(GROUP.+)\n</code> Reemplazar: <code>\$1\nSEX: \n</code>
4. Marcar el inicio de los discursos (SPEECH)	Buscar: <code>(SEX: \n)(\p{L})</code> Reemplazar: <code>\$1SPEECH: \$2</code>
5. Marcar el inicio de discurso en los discursos escritos (“in writing”)	Buscar: <code>(<omit>written</omit>)\n</code> Reemplazar: <code>\$1\nSPEECH:</code>
6. Colocar <i>header</i> antes del inicio del discurso	Buscar: <code>\n\n(NAME.+?)\n(GROUP POSITION.+?)\n(SEX:.+?)\n</code> Reemplazar: <code>\n\n<text>“\$1” “\$2” “\$3”</text>\n\$4</code>
7. Marcar el final del discurso (para facilitar análisis en Wordsmith)	Buscar: <code>(\n\nNAME)</code> Reemplazar: <code>#END_OF_SPEECH\$1</code>

Figura 3. Ejemplos de expresiones regulares para el etiquetado automático

Tras aplicar las expresiones regulares se han corregido de forma manual algunos errores de formato. En dicha revisión se han añadido puntualmente algunas expresiones como pueden ser “<omit>cut</omit>” (cuando se interrumpe al orador / a la oradora), “<omit>other_lang</omit>” (cuando quien habla cambia de código lingüístico y que se pueden omitir al generar los datos) o “[...]” (cuando se produce alguna omisión o, en el caso de las interpretaciones simultáneas, el audio resulta ininteligible). Teniendo en cuenta todos estos elementos, nuestro corpus ya está listo para introducirlo en el software de análisis, con una apariencia, ya etiquetado, parecida a la siguiente:

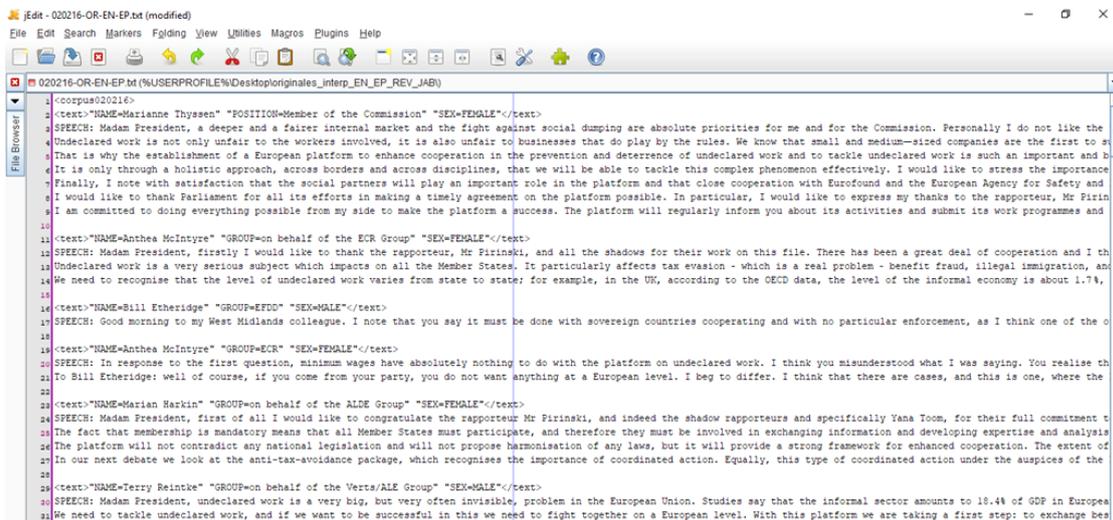


Figura 4. Ejemplo de discursos etiquetados documentalmente en XML

3.4. Herramientas de análisis: WordSmith Tools 6.0.

Una vez concluido el proceso de compilación y etiquetado de los textos, pasamos al análisis; y, para ello, resulta imprescindible contar con un buen *software* de análisis. En palabras de Roberts (2006, p.69), «a corpus is only as good as the query system you have to consult it».

La herramienta que cuenta con un mayor número de adeptos y una de las más potentes, dado su gran nivel de funcionalidad (Baker, 2006, p.76), es *WordSmith Tools*¹⁰. Desarrollada por Mike Scott, WordSmith Tools «is an integrated suite of programs looking at how words behave in text» (Scott, 2015, p.2).

Esta es la herramienta que hemos utilizado en el análisis; en particular la penúltima versión, la 6.0., de 2015. Si bien uno de los inconvenientes de WordSmith Tools es que no es de libre acceso, en nuestro caso hemos contado con una licencia gracias al grupo ECPC, que trabaja fundamentalmente con WordSmith Tools en sus análisis, lo que nos ha permitido llevar a cabo el análisis del presente trabajo.

De entre las múltiples utilidades de WordSmith Tools, en palabras de Calzada Pérez (2012, p.91), se podría destacar que es muy eficiente en el análisis de los corpus monolingües (no siendo así con los paralelos, cuyo procesamiento no permite) codificados en XML. De este modo, WordSmith Tools nos permite obtener listados de palabras clave, concordancias, listas de frecuencia, etc. en una interfaz relativamente sencilla que facilita el análisis, que llevaremos a cabo en el apartado siguiente.

¹⁰ Manual de uso de la versión 6.0 disponible en: <http://lexically.net/downloads/version6/wordsmith6.pdf> [Consulta: 23/04/2018]

4. Análisis del corpus

Muchas disciplinas comparten hoy en día cierto deseo de analizar el discurso político, y este también es el caso de los ETIC. De hecho, como apunta Ädel (2010), el discurso político-institucional está muy representado en los corpus traductores, sobre todo porque es más accesible que otros tipos de discursos, dada la obligación de transparencia de las instituciones y dada la ingente actividad traductora en el seno de las mismas. Así, no resulta extraño que los ETIC se hayan interesado por el estudio de este tipo de discurso, si bien «few generally available corpora exist which contain exclusively genres of political discourse» (*ibíd.*: p.593).

Mucho se ha hablado también, por otro lado, de la metodología y el tipo de análisis que precisa el discurso político. Según Gabrielatos y Baker (2008, p.7), estos análisis se han realizado a menudo desde un enfoque mixto capaz de integrar la metodología de la LC, «that allows for a higher degree of objectivity» (*ibíd.*), y el «general framework for problem-oriented social research» (Baker et al., 2008, p.279) que proporciona el ACD. No en vano, la parte interpretativa cobra especial importancia, en la medida que «without interpretation, any study in corpus linguistics would be incomplete» (Thornbury, 2010, p.274). De este modo, en nuestro análisis seguiremos una progresión que irá de una metodología cuantitativa a una cualitativa, que comenzará con un listado de frecuencia, para llegar a palabras clave, concordancias y colocaciones que nos permitan arrojar luz sobre cómo se habla de y se describe a las personas refugiadas en los originales en inglés e interpretaciones de otras lenguas al inglés de intervenciones parlamentarias europeas.

4.1. Listas de frecuencia (*wordlist*)

Una *wordlist* es una lista de palabras ordenadas en función de su frecuencia de aparición en un corpus; si bien igualmente pueden ordenarse alfabéticamente. El de la frecuencia, de acuerdo con Baker (2006, p.47) y Stubbs (1996, p.107), es un elemento imprescindible en el análisis de corpus, pues el lenguaje se basa en reglas y patrones que establecen qué puede —y no puede— decirse y, por lo tanto, la libertad de elección de un término o palabra, aun siendo uno sólo, ya refleja un posicionamiento ideológico que necesita analizarse.

Con todo, no podemos olvidar que este análisis presenta limitaciones, pues «it can be helpful in determining the focus of a text» (*ibíd.* 2006, p.71) pero «care must be taken not to make presuppositions about the way words are actually used within it» (*ibíd.*: p.71). En cualquier caso, resulta evidente que un diputado del PPE no utilizará la palabra ‘refugee’ del mismo modo que una diputada de los Verdes/ALE. De ahí la necesidad de un etiquetado documental y de no limitarnos al análisis cuantitativo, que, a lo sumo, podrá proporcionarnos “indicios” sobre qué dirección hemos de tomar para obtener resultados, que no lograríamos sin un análisis de, por ejemplo, concordancias o palabras clave en contexto, que amplíen los horizontes de nuestras hipótesis. Si bien es cierto que en el discurso

político se ha hecho mucho hincapié en las palabras (Wilson, 2003), al fin y al cabo el contexto, en tanto que «personal mental construct» (Van Dijk, 2002, p.18), es el que traslada el mensaje.

En lo que respecta a nuestro corpus, hemos creado tres listas de frecuencias: una para las versiones originales en inglés, otra para las interpretaciones al inglés y otra que integra ambas. No obstante, merece la pena comenzar analizando algunos datos estadísticos (v. **Figura 5**) sobre, por ejemplo, la variación léxica presente en originales e interpretaciones, que se percibe gracias a la *type/token ratio*. De acuerdo con Baker (2006, p.52), la *type/token ratio* es especialmente útil porque nos indica el grado de especificidad o complejidad lingüística del corpus. Resulta curioso ver un *Standardised Type/Token Ratio* mayor en las interpretaciones (40,37) que en los originales (38,27), lo cual viene a sugerir que la variación léxica de las primera es mayor que la de estos últimos, lo cual resulta paradójico ya que, de acuerdo con Baker (*ibíd.*), los textos escritos suelen presentar mayor variación léxica y los orales suelen ser más repetitivos. En cualquier caso, estos datos podrían ser indicio, por ejemplo, de mayor flexibilidad y agilidad mentales, de mayor precisión léxica o, quizá, de que las intérpretes adoptaron una postura menos prescriptivista que quienes transcribieron los discursos originales. Conviene indicar, no obstante, que para poder apuntalar estos datos e incluso formular hipótesis sólidas necesitamos un estudio más detenido que se aleja de los objetivos del presente trabajo.

Originales en inglés		Interpretaciones al inglés		Originales e interpretaciones	
file size	4.268.612	file size	692.794	file size	5.309.013
tokens (in text)	340.209	tokens (in text)	58.732	tokens (in text)	398.941
tokens used for word list	337.869	tokens used for word list	58.407	tokens used for word list	396.276
types	10.644	types	4.830	Types	11.674
type/token ratio (TTR)	3,15	type/token ratio (TTR)	8,27	type/token ratio (TTR)	2,95
standardised TTR	38,27	standardised TTR	40,37	standardised TTR	38,57
STTR std. dev.	59,53	STTR std. dev.	57,90	STTR std. dev.	59,82
STTR basis	1.000	STTR basis	1.000	STTR basis	1.000
mean word length (in characters)	4,89	mean word length (in characters)	4,61	mean word length (in characters)	4,85
word length std. dev.	2,77	word length std. dev.	2,56	word length std. dev.	2,75
sentences	13.375	sentences	2.860	Sentences	16.235
mean (in words)	50,09	mean (in words)	20,42	mean (in words)	44,86
std. dev.	2.030,07	std. dev.	12,46	std. dev.	1.842,59

Figura 5. Datos estadísticos sobre el corpus (Fuente: elaboración propia)

Antes de crear las listas de frecuencias, hemos decidido utilizar una *stop list* (fichero con extensión .stp), esto es, una lista de palabras que queremos dejar fuera del cómputo. Esta decisión se debe a que, en general, las palabras gramaticales suelen ocupar las primeras posiciones de las listas de frecuencias y para poder obtener resultados más significativos hemos decidido prescindir de algunas —

no de todas— las palabras gramaticales. Se ha usado la *stop list* que WordSmith Tools trae por defecto (*Basic English Stoplist*), formada por alrededor de 60 palabras. Con todo, de esta lista por defecto se han las palabras *the* y *that*, con tal de tenerlas en cuenta posteriormente en nuestros análisis.

N	Word	Freq.	%	Te
1	THE	22.790	6,62	
2	THAT	5.023	1,46	
3	END	2.196	0,64	
4	EU	2.072	0,60	
5	EUROPEAN	1.774	0,51	
6	REPORT	1.221	0,35	
7	OUR	1.200	0,35	
8	VOTED	1.096	0,32	
9	ALSO	1.064	0,31	
10	COMMISSION	993	0,29	
11	UNION	923	0,27	
12	MEMBER	905	0,26	
13	SHOULD	845	0,25	
14	AGAINST	835	0,24	
15	STATES	786	0,23	
16	MORE	780	0,23	
17	PRESIDENT	738	0,21	
18	DO	735	0,21	
19	VERY	663	0,19	
20	NEED	658	0,19	
21	PARLIAMENT	652	0,19	
22	SUPPORT	648	0,19	
23	AGREEMENT	631	0,18	
24	EUROPE	584	0,17	
25	FAVOUR	568	0,16	
26	THESE	560	0,16	
27	NEW	546	0,16	
28	IMPORTANT	540	0,16	

Figura 6. Wordlist de los originales

N	Word	Freq.	%	Te
1	THE	3.223	5,44	
2	THAT	1.156	1,95	
3	EUROPEAN	368	0,62	
4	REFUGEES	295	0,50	
5	WOMEN	284	0,48	
6	OUR	280	0,47	
7	NEED	267	0,45	
8	END	260	0,44	
9	PEOPLE	258	0,44	
10	EUROPE	235	0,40	
11	THANK	204	0,34	
12	VERY	201	0,34	
13	DO	198	0,33	
14	SHOULD	194	0,33	
15	MEMBER	187	0,32	
16	STATES	187	0,32	
17	ABOUT	178	0,30	
18	THESE	178	0,30	
19	TURKEY	172	0,29	
20	UNION	167	0,28	
21	PRESIDENT	154	0,26	
22	BORDERS	148	0,25	
23	ALSO	145	0,24	
24	CRISIS	144	0,24	
25	MORE	140	0,24	
26	COUNTRIES	139	0,23	
27	EU	139	0,23	
28	THINK	138	0,23	

Figura 7. Wordlist de las interpretaciones

Como se aprecia en las figuras anteriores (v. **Figura 6** y **Figura 7**), la palabra más frecuente tanto en los originales como en las interpretaciones es el artículo *the*, algo con lo que coinciden otros estudios (cf. Calzada Pérez, 2012; Calzada Pérez, 2017). Por otro lado, cabe señalar que el uso de *that* es más habitual en las interpretaciones que en los originales. Recordemos que Baker y Olohan (2000), de hecho, estudiaron el uso explícito de *that* en los textos traducidos y observaron que es más habitual que en los originales. Podemos quizá deducir que estamos ante el mismo fenómeno, aunque no podemos darlo por supuesto, ya que *that* tiene usos diversos. De nuevo, parece claro que se requiere investigación futura en este sentido.

Tampoco es de extrañar que *European* aparezca en las primeras posiciones de la lista de originales y de interpretaciones, pero resulta interesante que en los originales sea más frecuente la abreviatura (EU) que en las interpretaciones, lo que podría señalar cierta tendencia de las intérpretes a decir *European Union* en vez de *EU* o *the EU*. Por lo demás, tanto en una lista como en la otra son bastante frecuentes los verbos directivos, que buscan orientar la acción de la UE (*should, must, do, think, thank, help...*);

pero sobre todo en la *wordlist* de las interpretaciones, en la que queda reflejado el debate sobre las fronteras, los migrantes y solicitantes de asilo y el espacio Schengen, ya que la palabra *refugees* es la tercera más frecuente en el corpus, por encima de *women* pese a que uno de los dos debates que incluye el corpus tuvo lugar en el día internacional de la mujer.

De igual modo, tanto en una lista como en otra, encontramos muchas referencias internas a instituciones de la UE (*Commission, Parliament, Member States, Council*), o a la presidencia (*President* o *Chair*), siendo la primera mucho más frecuente en líneas generales, aunque en las interpretaciones también se ha usado *Chair*). Son habituales las referencias anafóricas, sobre todo con el uso de *our* o *them*, quizá en referencia a la UE (sería necesario comprobar qué viene tras *our*) y a las personas refugiadas (convendría observar en cotexto y contexto a qué se refiere *them*), y las espacio-temporales (*here, now, today...*). Esto queda patente en la *wordlist* de interpretaciones, donde encontramos todos los elementos necesarios para saber que se está hablando del acuerdo con Turquía para la devolución de personas refugiadas (*these, people, Turkey, agreement, refugees, borders...*).

Abundan las referencias al contexto (*crisis, problem, situation, issue...*), o a la búsqueda de soluciones (*solution, solidarity, help, support, protection*), fundamentalmente en las interpretaciones. Esto puede deberse a que las intérpretes utilizan más los sustantivos o a que las oradoras que se expresan en inglés hablan menos de soluciones frente al resto. Asimismo, son habituales las referencias a Grecia y a Turquía, que únicamente aparecen en la lista de las interpretaciones y que llegan a través de la interpretación a las personas que no entienden sino inglés. Por último, en nuestras listas aparecen palabras como *migrants, refugee/s, asylum [seekers], immigrants, children, women...*, que consultaremos con mayor detalle para ver cuál es el uso que se da a estas palabras, si hay confusión o no entre ellas, y si hay colocaciones o agrupaciones que destaquen.

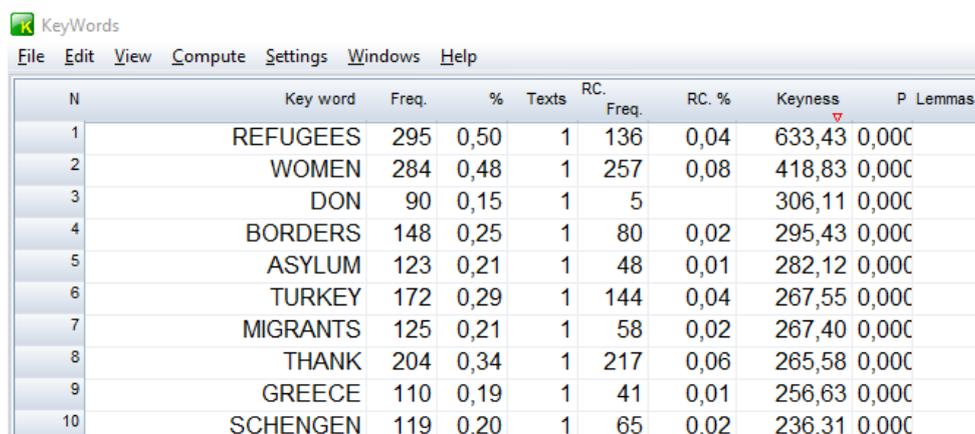
Como comentábamos en la introducción a este apartado, esta fase estadística nos proporciona numerosos indicios y nos da pistas sobre dónde buscar, pero no certezas suficientes en tanto en cuanto las palabras no están en contexto. Sin embargo, antes de adentrarnos en el listado de concordancias, que, de acuerdo con Marco (2013, p.19), es sin dudas el principal instrumento con el que trabaja quien centra su actividad en los corpus, creemos necesario analizar primero las palabras clave y los enlaces que se establecen entre ellas, con tal de obtener más “indicios” a partir de los que buscar.

4.2. Palabras clave (*keywords*)

Olohan (2004, p.81) señala que una lista de palabras clave consiste en una extensión de la lista de frecuencia. Para su obtención, WordSmith Tools compara las listas de frecuencia de dos corpus (uno de mayor tamaño, usado de referencia, y uno más pequeño, especializado) de modo que se obtiene una lista de palabras que son relativamente frecuentes en el pequeño pero que no están muy representadas en el de mayor tamaño.

Por su parte, Evison (2010, p.127) habla de *positive* y *negative* keywords, esto es, palabras que destacan por ser muy frecuentes (*positive*) o por ser infrecuentes (*negative*), que WordSmith Tools normaliza automáticamente; o sea, contrapone las frecuencias obtenidas de los corpus de distinto tamaño que se comparan para poder posteriormente expresar el resultado de acuerdo a un factor común. En este caso, se han comparado un corpus de intervenciones parlamentarias en inglés original correspondientes a los días de plenario de febrero y marzo de 2016 (de mayor tamaño, usado como referencia) y uno de menor tamaño compuesto de transcripciones de interpretaciones de otras lenguas al inglés de un día de febrero y un día de marzo.

Así, como se aprecia en las **Figura 8** y **Figura 9**, las palabras clave nos ayudan a confirmar parcialmente algunas hipótesis que surgen en la lista de frecuencia, en la medida en que palabras como *refugees* y *women* destacan por ser relativamente frecuentes, tanto en la lista de frecuencia como en la de palabras clave. Es el caso de *borders*, *asylum* o *migrants* e igualmente el de *Turkey* y *Greece*, que son los países que están más presentes en el contexto de los debates recogidos. Otra palabra que nos señala como clave es *don*, que en realidad se refiere a *don't*, y cuya presencia se debe a que en las transcripciones de las interpretaciones se ha usado la forma exactamente utilizada por quien interpretaba, mientras que en las versiones originales se ha usado la forma *do not* de manera uniforme.



N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P Lemmas
1	REFUGEES	295	0,50	1	136	0,04	633,43	0,00C
2	WOMEN	284	0,48	1	257	0,08	418,83	0,00C
3	DON	90	0,15	1	5		306,11	0,00C
4	BORDERS	148	0,25	1	80	0,02	295,43	0,00C
5	ASYLUM	123	0,21	1	48	0,01	282,12	0,00C
6	TURKEY	172	0,29	1	144	0,04	267,55	0,00C
7	MIGRANTS	125	0,21	1	58	0,02	267,40	0,00C
8	THANK	204	0,34	1	217	0,06	265,58	0,00C
9	GREECE	110	0,19	1	41	0,01	256,63	0,00C
10	SCHENGEN	119	0,20	1	65	0,02	236,31	0,00C

Figura 8. Lista de *positive keywords*



N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P Lemmas
1	EU	139	0,23	1	2.072	0,61	-158,17	0,00000000
2	REPORT	52	0,09	1	1.221	0,36	-156,36	0,00000000
3	FAVOUR	6	0,01	1	568	0,17	-138,67	0,00000000
4	THE	3.223	5,44	1	22.783	6,70	-136,47	0,00000000
5	UK	4		1	427	0,13	-106,99	0,00000000
6	AGREEMENT	23	0,04	1	631	0,19	-91,27	0,00000000
7	MARKET	7	0,01	1	415	0,12	-88,72	0,00000000
8	LEGISLATION	7	0,01	1	355	0,10	-71,61	0,00000000
9	DATA	3		1	284	0,08	-69,31	0,00000000
10	AGAINST	53	0,09	1	835	0,25	-68,91	0,00000000

Figura 9. Lista de *negative keywords*

Resulta interesante analizar las *negative keywords*, pues nos permiten darnos cuenta de que palabras como *EU*, *legislation* o *agreement* no son tan frecuentes en el corpus de interpretaciones mientras que son muy frecuentes en el de originales. Así, existirán ideas y conceptos que nos llegarán primordialmente a través de las interpretaciones o de los originales, lo que representa cierta pérdida (en cualesquiera de los casos). Lo mismo ocurre con *UK* pues no podemos olvidar que el corpus de interpretaciones comprende dos meses del año en que Reino Unido votó en referéndum su salida de la UE, de ahí que su representación en los discursos tenga mayor cabida.

Otra de las funciones de WordSmith Tools es que nos permite establecer enlaces entre palabras clave. O entre las palabras clave que selecciona y otras cuya *keyness* es inferior. Estos enlaces (*links*) son una forma de comenzar a agrupar las palabras y saber por dónde comenzar a buscar a la hora de analizar las concordancias. Por ejemplo, como vemos en la **Figura 10**, la palabra clave *refugee* suele ir ligada a palabras como *women*, *fleeing*, *people*, *Turkey* o *them*, siendo esta última especialmente relevante porque marca un distanciamiento de los oradores (o intérpretes) respecto de las personas refugiadas: ellas frente a nosotras, las europeas; y ese es el tipo de discurso que emite una UE deshumanizada. Por su parte, la palabra *women* está ligada a *violence* o *refugees*, quizá en referencia a los acontecimientos que tuvieron lugar en Colonia en diciembre de 2015; o la palabra *Turkey*, que ya lleva aparejada, por ejemplo, la palabra *billion*, en referencia a los tres mil millones de euros que la UE se comprometió a pagar a Turquía con tal de frenar la llegada de personas a Europa.

N	Key word	Link types	Hits	Linked KWS
1	REFUGEES	8	295	them, women, countries, fleeing, these, people, turkey, about
2	WOMEN	10	284	seekers, refugees, arriving, violence, need, people, Europe, these, day, about
3	TURKEY	4	172	billion, now, Greece, refugees
4	MIGRANTS	2	125	people, Greece
5	GREECE	2	110	Turkey, migrants

Figura 10. Enlaces entre palabras clave

4.3. Concordancias (*concord*)

No cabe duda de que el análisis de concordancias supone dar un salto cualitativo en cualquier investigación con corpus. De acuerdo con Sinclair (1991, p.32):

A concordance is a collection of the occurrences of a word-form, each in its own textual environment. In its simplest form it is an index. Each word-form is indexed and a reference is given to the place of occurrence in a text.

En definitiva, una concordancia es una lista de todas las ocurrencias que se obtienen al buscar una palabra particular en nuestro corpus y que aparecen rodeadas del contexto del que forman parte. Una de las ventajas de esta disposición, según Tribble (2010, p.175) y Baker (2006, p.77), es que nos permite

trabajar de forma ordenada, y tal disposición puede ir reordenándose según nuestros criterios (podemos usar comodines, buscar palabras a la izquierda y palabras a la derecha, ordenar de forma ascendente o descendente, etc.). Sin embargo, cabe tener en cuenta que la cantidad de palabras clave en contexto (KWIC) a veces puede ser muy alta y si quien investiga no tiene una idea clara de qué quiere buscar, es difícil obtener resultados significativos de las concordancias. De ahí la necesidad de un primer análisis cuantitativo que nos ahorre tiempo y nos ayude a gestionar los datos. Por lo tanto, nuestra búsqueda y análisis de concordancias se ha realizado partiendo de los resultados obtenidos en la lista de frecuencias y lista de palabras clave.

A título ilustrativo (**Figura 11**), buscamos una de las palabras clave en el corpus usando comodines (*refugee?*) y obtenemos una lista de concordancias, de la que conservamos las 50 primeras que nos servirán para comenzar un análisis mínimo sobre cómo se representa a las personas refugiadas, en este caso, en las interpretaciones. Obsérvese que análisis posteriores han de confirmar, refutar y siempre completar nuestros datos:

1 is year, to date, 138 000 **refugees** and migrants have arrived
2 Spain has taken in 18 **refugees** to date. And in months cut
3 say that 45% of the Afghan **refugees** arriving in Greece where
4 question of violence against **refugees** is very important so as
5 and discrimination against **refugees** in Europe have causes, and
6 European mechanism for aiding **refugees** in Turkey is duping us.
7 ought to stop migrants and **refugees** arriving. I will very much
8 that proposed shooting at **refugees**, we think that is not the
9 here is confusion between **refugees** and migrants since we lack
10 there are fights between **refugees** because of differences in
11 committed by the so-called **refugees** against our own people.
12 against refugees, Christian **refugees**, by other immigrants, in
13 and those who are economic **refugees**. Also, minors should not
14 not only to protect female **refugees**, but also to strengthen
15 daunting challenge of filtering **refugees**. When asylum seekers seek
16 to improve conditions for **refugees** and give protection in Jordan
17 providing protection for **refugees**. Thank you, President. Mr
18 safe and legal routes for **refugees** and asylum seekers is the
19 safe and legal ways for **refugees** and migrants, and, at the
20 do more to help genuine **refugees** and offer protection to
21 grants; these are genuine **refugees**, people to whom we can and
22 choose whether to help **refugees** or not a lot of people in
23 Member States need to help **refugees** and not have to open the
24 Fascist groups are humiliating **refugees**. We need a Europe which
25 countries that have taken in **refugees** and we must establish our
26 efficient way of taking in **refugees** from the region; not to
27 to resettle individual **refugees**, whereas we would need tens
28 been paid money to keep **refugees**. They got 3 billion, now
29 do not want to take more **refugees** for the European Union that
30 who come to the EU are not **refugees**, but are economic migrant
31 coming to Europe are not **refugees** and that we need to make
32 for more massive flows of **refugees**. How can you say that
33 the difficult situation of **refugees** massively fleeing conflict
34 about the invasion of **refugees**. There were groups of people
35 that uncontrolled influx of **refugees** into Europe must finish.
36 this case, the challenge of **refugees**. We need solidarity, but

37 can tackle the issue of **refugees**; it's not an Italian problem
 38 considered as political **refugees**, which by the way is the
 39 and those that are real **refugees**. Once again, you seem to have
 40 longer be able to receive **refugees**, in Lebanon we have already
 41 some 2.6 million registered **refugees**. The vast majority of the
 42 that Greece should return **refugees** back to Turkey, where the
 43 situations, especially Somali **refugees**. In many of the protracted
 44 the admission of Syrian **refugees** so that they do not have
 45 into centres for Syrian **refugees** in Turkey, for health, education
 46 and other assistance to **refugees** and host communities. This
 47 want to close the door to **refugees**. I think we need to remember
 48 around the world to treat **refugees** according to principles and
 49 struggling to cope with **refugees** from Syria, is not merely
 50 European strategy to deal with **refugees**. They might show up in

Figura 11. Concordancias obtenidas a partir de la búsqueda "refugee?" en el corpus de interpretaciones

Ateniéndonos al modelo de análisis propuesto por Baker (2006), Gabrielatos y Baker (2008) y Baker et al. (2008), encontrarnos, en primer lugar, una cuantificación evidente de las personas refugiadas; a veces de forma exacta (*138 000 refugees and migrants, 18 refugees, 2.6 million registered refugees*), y a veces de forma imprecisa (*massive flows, waves of, uncontrolled influx...*). También se describe a las personas refugiadas por su género o edad (*women refugees, children refugees...*), nacionalidad (*Syrian, Afghan, Somali...*) o incluso por su religión (*Christian refugees*). Por otro lado, se las clasifica de acuerdo con etiquetas que no se ajustan a la definición de lo que es una persona refugiada, cuando, por ejemplo, se habla de *political refugees* o de *economic refugees*; o incluso se llega a cuestionar en cierto modo su estatus, pues se habla de *real refugees* o *genuine refugees*, que, como revela nuestro etiquetado, pertenecen a miembros de grupos de la derecha y ultraderecha xenófoba. La deshumanización de las personas refugiadas llega hasta el punto de hablar de *filtering refugees*.

Abundan las referencias al hecho de que las personas refugiadas representan un *challenge, problem, difficult situation, threat, issue* para la Unión Europea, acompañadas de verbos como *cope with, deal with* o *tackle*, en referencia a la necesidad de gestionar la "crisis" o simplemente a devolver o reubicar personas refugiadas (*return, resettle*) personas refugiadas. Conviene señalar, asimismo, que, tanto en originales como interpretaciones, se observa cierta tendencia a buscar ayuda para las personas refugiadas o al menos a plantear la necesidad de tal ayuda, tanto desde un punto de vista general (*take in, keep, receive, aiding, providing protections, improving conditions*) como durante el trayecto hasta Europa (*safe and legal routes, safe and legal ways...*). No se observan, sin embargo, referencias a las, en palabras de Baker (2006, p.84), «tragic circumstances of refugees», o no al menos en sus países de origen, pues sí que encontramos palabras como *violence, discrimination* o *humiliating* en Europa. Finalmente hallamos la expresión *invasion of refugees*, pronunciada por una diputada del grupo GUE/NGL reproduciendo unas palabras pronunciadas por el entonces primer ministro danés contra la recepción de personas en Europa.

Resulta interesante, por otra parte, analizar el uso de los términos *immigrant(s)* y *asylum seeker(s)* en nuestro corpus, con tal de ver con qué frecuencia y cómo se usan o si hay confusiones entre términos (Figura 12). Como vemos, el término *immigrant* es más habitual en el corpus de interpretaciones que en el de originales, pese a ser este más grande, lo que quizá podría deberse a la repetición de la intérprete o bien a una visibilización del término mediante la interpretación. Lo mismo sucede con *migrant* y con *asylum seekers*. Este hecho no es sorprendente pues el corpus de interpretaciones está especializado en esta cuestión. Sin embargo, la palabra *migrant* es más frecuente en el corpus que *immigrant*.

Al igual que sucedía con *refugees*, también con *immigrants* se produce una cuantificación (*two million*) exacta e inexacta (*a massive influx of*) y una calificación, en absoluto ajustada a la definición de persona inmigrante pero recurrente en el corpus, que es *illegal immigrant(s)*. Además de mencionarse su condición, también se menciona su procedencia (*Middle East and North Africa*, por ejemplo), o bien se les trata como delincuentes (*a crime perpetrated only by immigrants*). Asimismo, se habla del movimiento de estas personas (*arrival, returning, channelled through...*).

1	Member States to accept	immigrant	quotas. The fact that	020216_10-INTER-EN-EP
2	take on migrants and	immigrants	and not expect Greece to	020216_10-INTER-EN-EP
3	Commission that there are	immigrants	who do not have the	020216_10-INTER-EN-EP
4	a crime perpetrated only by	immigrants.	Sadly, millions of women	030216-OR-EN-EP
5	For years and decades,	immigrants	from the Middle East and	030216-OR-EN-EP
6	the following: for every	immigrant,	illegally channelled through	020216_10-INTER-EN-EP
7	you want to help illegal	immigrants	at any costs before our citizens,	020216_10-INTER-EN-EP
8	claims and returning illegal	immigrants.	Is this so difficult when	020216-OR-EN-EP
9	agencies, pushing illegal	immigrants	to Greece. Unsurprisingly	020216_10-INTER-EN-EP
10	massive arrival of illegal	immigrants,	who most of them arrive	020216_10-INTER-EN-EP
11	Then the two million	immigrants	that have arrived, the several	020216_10-INTER-EN-EP
12	that it was mostly	immigrant	men that mugged, sexually harassed	030216-OR-EN-EP
13	swapping one set of	immigrants	for another, would you not agree?	090316-OR-EN-EP
14	the massive influx of	immigrants	we saw last year. We are closing	020216_10-INTER-EN-EP
15	extreme right on	immigrants,	which could lead to an unstoppable	030216-OR-EN-EP

Figura 12. Concordancia a partir de la búsqueda "immigrant?" en el corpus

1	movement only for terrorists,	asylum seekers	and kalashnikovs.
2	traffickers, and dealing with	asylum seekers,	I said. And I expect on
3	over 440 000 refugees and.	asylum seekers	in neighbouring countries
4	management of the flow of	asylum seekers	into Europe is a stress test
5	on women refugees and	asylum seekers	is about bringing a much needed
6	legal routes for refugees and	asylum seekers	is the only way to prevent
7	more than almost 200000	asylum seekers	last year and it had to borrow
8	refugees and	asylum seekers,	LGBTI persons, or persons with disabilities.
9	these women refugee.	asylum seekers	So we want to be realistic
10	need to protect thereby	asylum seekers,	respecting our international
11	female and LGBTI	asylum seekers,	as they are victims of extra violence

12	massive influx of migrants and	fortune seekers	can be combined with
13	the way that these	fortune seekers	deal with our children and women
14	improve the lives of	job-seekers	across Europe, and especially for
15	since today too many	job seekers	and employers are simply not aware

Figura 13. Concordancia obtenida a partir de la búsqueda "seeker?"

Si comparamos los resultados obtenidos en la figura 12 con los de la **Figura 13**, que se corresponden a la concordancia obtenida a partir de la búsqueda *seeker?* en el corpus, obtenemos resultados similares, en la medida en que también se produce una cuantificación o se hace referencia a la gestión de los flujos (*flows*). El término más frecuente es *asylum seekers*, que a menudo suele ir acompañado, como vemos en esta muestra, de *refugees*, *women refugees*, *migrants* o *LGBTI*, corriéndose el riesgo de vulnerabilizar al colectivo. De nuevo se les trata como delincuentes (véase la primera o la decimotercera concordancias de la tabla) o se habla de la necesidad de ofrecerles ayuda (*improve the lives of, protect*), o, sencillamente, de buscar soluciones (*deal with, management...*). Otra colocación, no tan habitual pero preocupante, es *fortune seekers*, que, al igual que sucedía con *illegal immigrant*, se aleja totalmente de aquello que se entiende por persona solicitante de asilo y se enmarca en un discurso ultraconservador, xenófobo y populista.

4.4. Colocaciones (*collocations*) y agrupaciones (*clusters*)

Como reza la famosa frase de Firth (1957): «You shall know a lot about a word from the company that it keeps». Y es que, en palabras de Baker (2006, p.94-95), una colocación es una palabra que aparece regularmente al lado de otra y cuya relación es significativa desde un punto de vista estadístico. De este modo, identificar patrones de agrupación de palabras nos permite describir e identificar discursos y resumir las relaciones entre las palabras más importantes de un corpus.

Así, siguiendo la tendencia observada en las listas de frecuencia, una de las palabras que frecuentes en el corpus es *crisis*, por lo que a continuación analizamos (**Figura 14**) cuáles son las colocaciones y agrupaciones más representativas que se dan con esta palabra en el corpus de originales y en el de interpretaciones.

ORIGINALES			INTERPRETACIONES		
N	Cluster	freq.	N	cluster	freq.
1	refugee crisis	46	1	refugee crisis	23
2	humanitarian crisis	19	2	migration crisis	13
3	financial crisis	19	3	deal with this crisis	6
4	military and humanitarian crisis	17	4	migrant crisis	5
5	ongoing military crisis	17	5	deal with crisis	5
6	migration crisis	16	6	refugee crisis in	4
7	current migration crisis	10	7	tackle this crisis	4
8	migration and refugee crisis	9	8	response to the crisis	4
9	the ongoing crisis	9	9	the European Union crisis	3
10	dealing with the crisis	7	10	this huge crisis	3
11	solution to the crisis	7	11	facing the crisis	2
12	response to the crisis	6	12	humanitarian crisis	2

13	the political crisis	4	13	crisis of the borders	1
14	a crisis situation	3	14	financial crisis	1
15	serious humanitarian crisis	3	15	crisis of the Union	1

Figura 14. Agrupaciones a partir de *crisis*

A partir de estos datos, resulta evidente que la expresión “refugee crisis” ha tenido hondo calado en los discursos del PE, en la medida en que, tanto en originales como interpretaciones en y al inglés, es la agrupación más habitual. Junto con eso, es frecuente la especificación del tipo de crisis, aunque, en el caso de los originales, se hace más hincapié en la idea de crisis humanitaria y en el de interpretaciones, en la idea de hacer frente a la crisis (*deal with, tackle, response to, facing...*). Como vemos, ambos corpus nos aportan información distinta. Sin interpretación, como ya observa Calzada Pérez para la traducción (2017), se pierde gran parte del discurso. Porque no podemos olvidar que, con la desaparición de la traducción, la única forma posible de hacer llegar los discursos es mediante la interpretación.

Sólo en el caso de los originales en inglés la crisis política forma un *cluster*, aunque, no en vano, sólo en las interpretaciones al inglés se habla de una crisis de la UE, lo cual significa que, de febrero a marzo de 2016, la situación política pudo cambiar y dejar de percibirse como una situación de simple crisis política para pasar a considerarse una crisis de la UE. En todo caso, la idea de crisis y la necesidad de encontrar soluciones atraviesan ambos discursos. Esta crisis aparece también cuantificada y situada en espacio y tiempo (*huge crisis, ongoing crisis, crisis of the borders...*).

Ahora bien, y a modo de conclusión, el reto se halla en entender por qué tanto en el inglés original como en el inglés interpretado de las intervenciones se habla de las personas refugiadas como una tragedia, el caos o una crisis que necesita afrontarse en vez de hablar sobre cómo subvertir los discursos fascistas y xenófobos camuflados de populismo que triunfan en el seno de la UE, de los que son víctimas las personas refugiadas, entre otras.

5. Conclusiones

El objetivo de este trabajo era explorar cómo se representa a las personas refugiadas en intervenciones parlamentarias de la UE originalmente en inglés e interpretaciones al inglés desde otras lenguas distintas, partiendo de que estos discursos se desarrollan en el contexto espacial de la comunidad discursiva del PE, y en el contexto temporal de los meses de febrero y marzo de 2016. Para este análisis nos hemos servido de una metodología mixta y de los postulados de los ETIC, gracias a los cuales hemos llegado a la conclusión de que, a partir de las intervenciones analizadas en inglés original e interpretado, en el PE se asocia la realidad de las personas refugiadas a la gestión de una crisis que se presenta en un momento de confusión —en todos los sentidos—, a lo que se suma la desaparición de la traducción, con la consiguiente falta de transparencia sobre y accesibilidad a los asuntos que se tratan en la UE.

De ahí que, frente a esta nueva situación, la interpretación cobre una importancia fundamental para cumplir las obligaciones de transparencia y para garantizar la recepción de discursos (tanto de los positivos como de los negativos). De este modo, y parafraseando la célebre frase de Steiner: «Without interpreting we would inhabit parishes bordering on silence»; sin interpretación, la única lengua de Europa es el silencio, y el silencio nunca es para bien.

El uso de las transcripciones de las interpretaciones, además, nos ha ayudado a observar de forma más directa las ideas imperantes en la comunidad discursiva, y el análisis mixto de los datos obtenidos nos ha permitido una categorización objetiva y flexible de los mismos. Por otra parte, la semiautomatización del proceso de compilación, etiquetado y análisis, aunque con limitaciones, nos ha permitido acelerar la obtención de unos resultados que, a nuestro entender, son todavía aplicables en la actualidad.

No obstante, resulta evidente que una mayor cantidad de textos, en otras lenguas, junto con los correspondientes etiquetados documental y gramatical, nos habrían permitido obtener resultados aún más representativos y, por lo tanto, más concluyentes sobre el objeto de nuestro estudio. Sin embargo, como hemos comprobado, el esfuerzo de compilar corpus, así como los límites de carácter espacio-temporal, constriñen a quien trabaja con corpus y obligan en muchas ocasiones a reducir la muestra o a que tal muestra deje de ser representativa para el momento en que se obtienen los primeros resultados.

A todo ello cabría añadir la necesidad de automatizar un proceso que supone tiempo, que es el de la transcripción o, al menos, el de una transcripción que refleje de forma fehaciente el marco multimodal en el que se desarrollan los discursos. Por otro lado, puesto que el período de tiempo que abarca nuestros discursos no permite tener en cuenta las traducciones, sería interesante comparar traducciones e interpretaciones de otros períodos para comprobar qué se ha ganado y qué se ha perdido. Finalmente, no

podemos pasar por alto que, con la salida definitiva de Reino Unido de la UE, el peso del inglés podría disminuir sensiblemente.

A modo de conclusión, conviene señalar que los ETIC, que la explotación de corpus en traducción e interpretación, aún se encuentran en pleno desarrollo. Y en este sentido, la inclusión de la interpretación sirve para aportar riqueza, luz e inclusión a nuestros análisis. No obstante, hemos de tener presentes las opiniones que, desde hace tiempo, invitan a hacer un uso prudente, realista y responsable de los corpus o simplemente a tener en cuenta sus limitaciones (cf. Malmkjær, 1998; Kenny, 2001, p.71) y peligros (Tymoczko, 1998); pues, como señala Borja (2008), no podemos olvidar que los corpus los construyen seres humanos.

Bibliografía

- ÄDEL, Annelie (2010). How to use corpus linguistics in the study of political discourse? En: Anne O'KEEFFE y Michael McCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 591-604). Nueva York: Routledge.
- ADOLPHS, Svenja y KNIGHT, Dawn (2010). Building a spoken corpus. What are the basics? En: Anne O'KEEFFE y Michael McCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 38-52). Nueva York: Routledge.
- ASTON, Guy (2018). Acquiring the Language of Interpreters: A Corpus-based Approach. En: Mariachiara RUSSO, Claudio BANDAZZOLI y Bart DEFRANCO (eds.), *Making Way in Corpus-based Interpreting Studies*. Berlín: Springer.
- BAKER, Mona (1993). Corpus Linguistics and Translation Studies: Implications and Applications. En: Mona BAKER, Elena TOGNINI BONELLI y Gill FRANCIS (eds.), *Text and technology. In honour of John Sinclair* (pp. 233-250). Ámsterdam: John Benjamins.
- BAKER, Mona y OLOHAN, Maeve (2000). Reporting That in Translated English: Evidence for Subconscious Processes of Explication? *Across Languages and Cultures*, 1(2), pp. 141-158.
- BAKER, Paul (2006). *Using Corpora in Discourse Analysis*. Londres: Continuum.
- BAKER, Paul, GABRIELATOS, Costas, KHOSRAVINIK, Majid, KRZYŻANOWSKI, Michael, McENERY, Tony y WODAK, Ruth (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273-306.
- BAUMAN, Zygmunt (2016). *Extraños llamando a la puerta*. Barcelona: Paidós. [Traducción: Albino Santos Mosquera].
- BAUMAN, Zygmunt (2017). *Retrotopía*. Barcelona: Paidós. [Traducción: Albino Santos Mosquera].
- BIBER, Douglas, CONRAD, Susan y REPPEN, Randi (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- BORJA ALBI, Anabel (2008). Corpora for Translators in Spain. The CDJ-GITRAD Corpus and the GENTT project. En: Gunilla ANDERMAN y Margaret ROGERS (eds.), *Incorporating Corpora: The Linguist and the Translator* (pp. 243-263). Clevedon: Multilingual Matters Ltd.
- CALZADA PÉREZ, María (2007). *Transitivity in Translating: The Interdependence of Texture and Context*. Preface by Ian Mason. Berlín: Peter Lang.
- CALZADA PÉREZ, María (2012). *Corpus comparables y paralelos de discursos parlamentarios europeos (ECPC): MaxiECPC*. Granada: Entorno Gráfico.
- CALZADA PÉREZ, María (2017). What is kept and what is lost without translation? A corpus-assisted discourse study of the European Parliament's original and translated English. *Perspectives*, 26(2), 277-291. <https://doi.org/10.1080/0907676X.2017.1359325>
- COMISIÓN EUROPEA (s. f.). Declaración UE-Turquía, 18 de marzo de 2016 - Consilium. Recuperado de: <http://www.consilium.europa.eu/es/press/press-releases/2016/03/18/eu-turkey-statement/>. [Consulta: 05/05/2018]
- EVISON, Jane (2010). What are the basics of analysing a corpus? En: Anne O'KEEFFE y Michael McCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 122-135). Nueva York: Routledge.

- FAIRCLOUGH, Norman (1989). *Language and Power*. Londres: Longman.
- FAIRCLOUGH, Normal (1993). *Discourse and Social Change*. Cambridge: Polity Press.
- GABRIELATOS, Costas y BAKER, Paul (2008). Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. *Journal of English Linguistics*, 36(1), 5-38. DOI: <https://doi.org/10.1177/0075424207311247>
- GRANGER, Sylvianne (2003). The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies. En: Sylvianne GRANGER, Jean LEROT y Stéphanie PETCH-TYSON (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* (pp. 17-30). Ámsterdam y Nueva York: Rodopi.
- KENNING, Marie-Madeleine (2010). What are parallel and comparable corpora and how can use them? En: Anne O'KEEFFE y Michael McCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics*. Nueva York: Routledge.
- KENNY, Dorothy (2001). *Lexis and Creativity in Translation: A Corpus-based Study*. Manchester: St. Jerome.
- KOSKINEN, Kaisa (2011). Institutional translation. En: *Handbook of Translation Studies* (Vol. 2, pp. 54-60). Londres y Nueva York: Routledge.
- KÜBLER, Natalie y ASTON, Guy (2010). Using corpora in translation. En: Anne O'KEEFFE y Michael McCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 501-515). Nueva York: Routledge.
- LAVIOSA, Sara (2002). *Corpus-based Translation Studies*. Ámsterdam y Nueva York: Rodopi.
- LEECH, Geoffrey (2004). *Adding Linguistic Annotation*. En: Martin WYNNE (ed.), *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxford Books.
- MARCO BORILLO, Josep (2013). Els estudis de traducció basats en corpus. En: Llum BRACHO LAPIEDRA (ed.), *El corpus COVALTT: un observatori de fraseologia traduïda*. Berlín: Shaker Verlag.
- MARTÍNEZ MARTÍNEZ, José Manuel y SERRAT ROOZEN, Iris (2012). ECPC: el discurso parlamentario europeo desde la perspectiva de los estudios traductológicos de corpus. *LinguaMÁTICA*, 4(2), 65-73.
- McENERY, Tony y HARDIE, Andrew (2012). *Corpus Linguistics. Method, Theory and Practice*. Cambridge: Cambridge University Press.
- McENERY, Tony y XIAO, Richard (2008). Parallel and Comparable Corpora: What is Happening? En: Gunilla ANDERMAN y Margaret ROGERS (eds.), *Incorporating Corpora: The Linguist and the Translator* (pp. 18-31). Clevedon: Multilingual Matters Ltd.
- O'KEEFFE, Anne y MCCARTHY, Michael (eds.). (2010). *The Routledge Handbook of Corpus Linguistics*. Nueva York: Routledge.
- OLOHAN, Maeve (2004). *Introducing Corpora in Translation Studies*. Londres y Nueva York: Routledge.
- RABADÁN ÁLVAREZ, Rosa y FERNÁNDEZ NISTAL, Purificación (2002). *La traducción inglés-español: fundamentos, herramientas y aplicaciones*. León: Universidad de León.
- REPPEN, Randi (2010). Building a corpus. What are the key considerations? En: Anne O'KEEFFE y Michael McCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 31-37). Nueva York: Routledge.
- ROBERTS, Roda P. (2006). Corpora and Translation. En: Lynne BOWKER (ed.), *Lexicography, Terminology, and Translation* (pp. 201-214). Ottawa: Ottawa University Press.

- SCOTT, Mike (2015). WordSmith Tools Manual. Recuperado de:
<http://lexically.net/downloads/version6/wordsmith6.pdf>. [Consulta: 05/05/2018]
- SINCLAIR, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SINCLAIR, John (2004). Corpus and Text-basic Principles. En: Martin WYNNE (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. (pp. 1-16). Oxford: Oxford Books.
- STUBBS, Michael (1996). *Text and Corpus Analysis*. Londres: Blackwell.
- SWALES, John M. (1990). *Genre Analysis. English in Academic and Research Setting*. Cambridge: Cambridge University Press.
- TEUBERT, Wolfgang y ČERMÁKOVÁ, Anna (2004). Directions in corpus linguistics. En: Michael A. K. HALLIDAY (Ed.), *Lexicology and Corpus*. Londres y Nueva York: Continuum.
- THORNBURY, Scott (2010). What can a corpus tell us about discourse? En: Anne O'KEEFFE y Michael MCCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 270-287). Nueva York: Routledge.
- TOGNINI BONELLI, Elena (2001). *Corpus linguistics at work*. Ámsterdam/Filadelfia: John Benjamins.
- TOGNINI BONELLI, Elena (2010). Theoretical overview of the evolution of corpus linguistics. En: Anne O'KEEFFE y Michael MCCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 14-27). Nueva York: Routledge.
- TRIBBLE, Christophe (2010). Concordances and how they are used. En: Anne O'KEEFFE y Michael MCCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 167-183). Nueva York: Routledge.
- TYMOCZKO, Maria (1998). Computerized Corpora and the Future of Translation Studies. *Meta: Journal Des Traducteurs*, 43(4), 652. <https://doi.org/10.7202/004515ar>
- VAN DIJK, Theun A. (1995). Discourse Analysis as Ideology Analysis. En: Christina SCHÄFFNER y Anita L. WENDEN (eds.), *Language and Peace* (pp. 17-33). Londres y Nueva York: Routledge.
- VAN DIJK, Theun. A. (2002). Political discourse and Ideology. En: Clara Ubaldina LORDAS y Montserrat RIBAS (eds.) (pp. 15-34). Barcelona: UPF.
- WILSON, John (2003). Political Discourse. En: Deborah SCHRIFFIN, Deborah TANNEN y Heidi E. HAMILTON (eds.), *The Handbook of Discourse Analysis*. Blackwell.